

## **CAT Tools e Tradução de Máquina, a revolução nas ferramentas de tradução à luz dos estudos da Linguística**

### **CAT Tools and Machine Translation, the revolution in translation tools in the light of Linguistics studies**

Luciana Galeani Boldorini<sup>1\*</sup>, Filomena Elaine Paiva Assolini<sup>1</sup>

---

#### **RESUMO**

Este artigo tenta traçar uma linha temporal breve sobre a trajetória da tradução entre idiomas, especialmente entre o português e o inglês, e culminar na Machine Translation (Tradução de Máquina) e, observar o comportamento das CAT Tools (Computer-Assisted Translation) como ferramentas aliadas dos tradutores contemporâneos, aumentando a produtividade e deixando a tradução/versão mais uniformes.

**Palavras-chave:** tradução; tradução automática; linguística computacional; CAT Tools.

---

#### **ABSTRACT**

This article attempts to trace a brief timeline on the trajectory of translation between languages, especially between Portuguese and English, and culminate in Machine Translation and observe the behavior of CAT Tools (Computer-Assisted Translation) as tools allies of contemporary translators, increasing productivity and making the translation/version more uniform.

**Keywords:** translation; automatic translation; computational linguistics; CAT Tools.

---

---

<sup>1</sup> Universidade de São Paulo

\*E-mail: [lucianagaleani@gmail.com](mailto:lucianagaleani@gmail.com)

## INTRODUÇÃO

*“Ever since computers were invented, we have wondered whether they might be made to learn. If we could understand how to program them to learn-to improve automatically with experience-the impact would be dramatic.”. (Mitchell, 1997)*

Se fossemos traçar toda a história da tradução de idiomas no mundo, não faltariam páginas repletas para desenhar esse percurso e seria um tanto quanto impreciso em datação e conteúdo. O que podemos assegurar é que a linguagem sempre esteve presente na história da humanidade, na formação das sociedades e implicando em como os seres humanos se relacionam.

Como bem exemplificado pelo professor José Luiz Fiorin (2008, p.29-53):

“A linguagem é onipresente na vida de todos os homens. Cerca-nos desde o despertar da consciência, ainda no berço; segue-nos durante toda a nossa vida, em todos os nossos atos, e acompanha-nos até na hora da morte. Sem ela, não se pode organizar o mundo do trabalho, pois é ela que permite a cooperação entre os seres humanos e a troca de informações e experiências. Sem ela, o homem não pode conhecer-se nem conhecer o mundo. Sem ela não se exerce a cidadania, porque ela possibilita influenciar e ser influenciado. Se ela não se pode aprender. Sem ela não se podem expressar sentimentos. Sem ela, não se podem imaginar outras realidades, construir utopias e sonhos. Sem ela não se pode falar do que é nem do que poderia ser.”

## UM BREVE HISTÓRICO

Desde que os seres humanos despertaram o interesse em desbravar territórios e descobrir o que havia além do mundo que conheciam e ir além de suas práticas sociais, a comunicação feita entre povos sofreu com a urgência da necessidade de traduzir para poder relacionar-se. Entender, comunicar-se e trocar cultura e experiências virou ponto central entre as relações humanas e a tradução teve seu início.

Bem como imaginamos, a tradução teve seu início oralmente (pela interpretação) por meio de gestos e imagens e só depois evoluiu para a escrita diacrônica para a narrativa de fatos e notícias. No Brasil, a exemplo, o primeiro contato com uma língua estrangeira, foi o português de nossos colonizadores. Foi-se necessária a intervenção dos chamados

línguas, que, mesclando-se aos índios nativos, descobriam meios da comunicação ser feita e assim serem realizados os escambos. Muitos *línguas* se tornaram famosos por ajudar nas relações de colonização e catequização dos índios nativos.

“Os mais famosos *línguas* desta época foram Afonso Ribeiro, João Ramalho e Diogo Álvares Correia (o Caramuru), que muito ajudaram na manutenção da relação de sobrevivência dos portugueses com o auxílio dos índios nas novas terras: João Ramalho auxiliou Martim Afonso na fundação da cidade de São Vicente em 1532 e Diogo Álvares ajudou Tomé de Souza na fundação da cidade de Salvador em 1549. Embora a História dê maior visibilidade a estes dois homens, convém mencionar que suas esposas - Catarina Álvares Paraguaçu (esposa de Diogo Álvares Correia) e Mbicy (Bartira, Burtira ou Isabel Dias – esposa de João Ramalho), ambas índias - talvez tenham sido as duas primeiras intérpretes/intermediárias do Brasil quinhentista, responsáveis por lhes ensinar a língua indígena e auxiliar a entrar em contato com a cultura indígena.” (REIS; MILTON, 2016, p.4)

Desde então, as técnicas de aquisição de idiomas e tradução têm evoluído muito. Foi-se a época que o tradutor se sentava com o texto e um dicionário e transcrevia todo o conteúdo para o papel, revisava e então podia fazer a plena compreensão do discurso.

Com os avanços tecnológicos caminhando a passos largos, a área de tradução pode colher benefícios desse feito. Hoje, com o advento da internet também, várias ferramentas e aplicativos foram desenvolvidos para nos auxiliar nessa tarefa.

É de conhecimento amplo e mundial o aplicativo *Google Tradutor*, que é a ferramenta mais conhecida por fazer traduções rápidas para vários idiomas e por ser de fácil acesso e uso. Insere-se uma palavra de um lado e esta é vertida ou traduzida.

“De acordo com SMAAL (2010), a intenção das empresas que desenvolvem ferramentas de tradução automática é diminuir a distância entre as línguas. Só o fato de o Google Tradutor ser disponibilizado gratuitamente e online facilita para que qualquer pessoa com acesso à internet tenha contato com ele, aumentando sua popularidade.” (CAPRIOLI; DANIEL, 2013, p.332)

O *Google Tradutor* utiliza-se de comparações entre idiomas (comparando termos, palavras e expressões entre definições de dicionários) e muito vocabulário foi agregado por informações que foram coletadas pelo próprio programa via documentos bilíngues.

“A ferramenta não é perfeita e autores como GOMES (2010) afirmam que, ao se dedicar mais tempo em sua utilização, é possível notar que a tradução é precisa em alguns pontos, entretanto, pode produzir sentenças confusas e malfeitas em alguns casos. Por desconhecerem esse fato, os usuários leigos se decepcionam ao verificarem que a

ferramenta não é tão perfeita quanto os desenvolvedores afirmam, os quais criam grande expectativa diante do texto traduzido, sendo que muitos não enfatizam a importância do aperfeiçoamento do trabalho humano de traduzir (ALFARO; DIAS 1998).” (CAPRIOLI; DANIEL, 2013, p.333)

## À LUZ DA LINGUÍSTICA

Sabemos que desde os primeiros estudos sobre a língua que, a grosso modo, as palavras não têm somente um significado e se formos mais além, como descrito por um dos primeiros estudiosos sobre linguística, o mestre genebrino Ferdinand de Saussure:

“Para Saussure, a língua não é nada mais que um sistema de valores puros. Descartando toda possibilidade de que a língua pudesse ser uma descrição do mundo, o linguista apresenta sua teoria, enfocando a língua como um fato social, produto da coletividade, que estabelece os valores desse sistema através da convenção social, sobre a qual o indivíduo não teria nenhum poder. Para perceber que a língua não é senão um sistema de valores, o linguista parte da análise de dois de seus aspectos básicos, as ideias e os sons.” (RODRIGUES, 2008, p.9)

Como citado pelo pai da linguística, se as ideias fazem parte dos aspectos básicos da língua, não há como se ter o mesmo valor de palavra quando temos ideias diferentes. Para Saussure o valor do signo só se dará, em suma, na presença de outros signos, que vão determinar o seu significado dentro de um conjunto de valores atribuídos.

“De um lado, o conceito nos aparece como a contraparte da imagem auditiva no interior do signo, e, de outro, este mesmo signo, isto é, a relação que une seus dois elementos, é também, e de igual modo, a contraparte dos outros signos da língua. (...) A língua [é] um sistema em que os termos são solidários e o valor de um resulta tão-somente da presença simultânea de outros (SAUSSURE, 2002, p. 133 IN: RODRIGUES, 2008, p.10).

Um mesmo signo pode ter diferentes significados de acordo com o contexto estabelecido, fatores socioculturais atribuídos e também a historicidade. As variações de significados são arbitrárias e não se seguem um padrão. A linguagem, como uma produção social, está sujeita a modificações decorrentes das interações sociais em que está inserida. Como exemplo, podemos ver um exemplo de como um termo pode sofrer alterações de significado de acordo com as interações sociais, retirado das aulas do mestre genebrino onde se tem:

“ (...) é fácil perceber que as línguas têm diferentes formas de nomear aspectos muitas vezes idênticos da realidade cotidiana. Um dos exemplos de Saussure é o termo *sheep*, do inglês. Esse termo equivale em significado, porém, não em valor, aos termos carneiro, do português e *mouton*, do francês. Ainda que os três sejam capazes de significar o

animal, em inglês se diz *mutton* para significar a porção de carne de carneiro já preparada para servir como alimento. Assim, o valor do termo inglês difere no interior da língua, não sendo, portanto, equivalente aos termos francês e português. ” (RODRIGUES,2008, p.11)

Assim, com os estudos aprofundados em linguística e suas vertentes, temos, com o advento da contemporaneidade e os avanços tecnológicos, o surgimento da Linguística Computacional, ou seja, o campo de pesquisa entre a linguística e a ciência computacional “tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural. ” (VIEIRA; LIMA, 2001)

Sendo assim, a Tradução Automática (doravante denominada T.A.), ou a tradução por auxílio de máquinas, ainda está longe de substituir as habilidades humanas, visto que máquinas (até então) não são dotadas de pensamentos e formas naturais de organização de linguagem.

Na psicogênese da linguagem, para existir a língua falada ou escrita, há de se existir um pensamento anteriormente. Este, através de ferramentas como o lápis, o papel e o computador, podem ser a ter a expressão do sujeito em sociedade. Sua linguagem escrita. Sua marca social.

Diante do desenvolvimento tecnológico da contemporaneidade, seria inconcebível imaginar o tradutor sentando em sua mesa de trabalho com uma folha de papel e um lápis e uma gama de dicionários à sua frente:

“Conforme similarmente expresso pela Tradutora Tamara Barile numa revista especializada no setor: [...] seria inconcebível ver, hoje, alguém trabalhando como trabalhávamos há 30 anos: usando o computador sem utilizar plenamente seus recursos (ou seja, utilizando-o como uma máquina de escrever), deixando de aproveitar plenamente o que a Internet tem a oferecer de bom ou, ainda, perdendo contatos com clientes e colegas por se recusar a utilizar o formato eletrônico de comunicação. [...] Todas essas mudanças exigem do tradutor um esforço constante de atualização e exigem que ele esteja aberto às novas formas de trabalho e de contatos profissionais.” (BARILE, 2006: 133-134 IN: ARAUJO, NOBRE, 2016, p.18)

Máquinas podem não serem dotadas de pensamentos, mas possuem uma habilidade muito interessante e importante que foi e é cada vez mais estudado e aprimorado nos estudos tecnológicos contemporâneos: o conceito de *Machine Learning* (Aprendizado de Máquina, doravante denominado A.M.).

O conceito de Aprendizado de Máquina (A.M.) surge com o desenvolvimento de estudos em Inteligência Artificial (I.A.) que “que visa ao desenvolvimento de programas de computador com a capacidade de aprender a executar uma dada tarefa com sua própria experiência. ” (FACELI et al., 2011). Ou, como diria o estudo referência dos autores Stuart Russel e Peter Norvig “Inteligência Artificial” (2013) onde os autores apontam a discussão filosófica sobre o poder das máquinas de pensarem, citam o grande cientista Alan Turin em seu ensaio *Computing Machinery and Intelligence* (Turing, 1950) onde o estudioso conseguiu “enganar” pessoas através de trocas de mensagens escritas, muito simplificada, fazendo as pessoas acreditarem que estavam falando com outra pessoa e não com uma máquina. Assim, Turin, há mais de 50 anos atrás, concebe os primeiros esboços do Aprendizado de Máquina e delimita que através do “argumento de inaptidão”, a máquina não poderá: “Ser amável, diligente, bonito, amigável, ter iniciativa, senso de humor, distinguir o certo do errado, cometer enganos, apaixonar-se, gostar de morangos e creme, fazer alguém se apaixonar por ela, aprender a partir da experiência, usar palavras corretamente, ser o sujeito de seu próprio pensamento, ter tanta diversidade de comportamento quanto o homem, fazer algo realmente novo.”

Como muito bem-conceituado por Alan Turin, habilidade das máquinas em aprender (*Machine Learning*) é de grande valia para processo e automações cotidianas atuais e, por ele frisado, a máquina, ainda, não “pode usar palavras corretamente”, o que só reforça que ainda, nada substitui ainda o conhecimento e aprendizado humano, mas, algumas ferramentas, munidas de Inteligência Artificial, podem ajudar e muito na tradução contemporânea. Hoje, mais do que nunca, o tradutor deve se valer dessas inovações tecnológicas para o dinamismo da comunicação global e facilitação de seu trabalho.

A *Machine Translation* (Tradução de Máquina), também chamada de *Automated Translation* (Tradução Automática, a T.A.) é “um subcampo da Linguística Computacional que investiga o uso de programas de computador que traduzem o texto e o discurso de um idioma para o outro”. (ERWEN; WENMING, 2013). Desde o início, a T.A. tem sido muito discutida entre estudiosos de vários campos por sua falta de precisão em atribuir um significado preciso à um signo dependo do contexto em que este está inserido. Como vimos anteriormente, o computador não é capaz de exprimir ou analisar sentimentos, como então seria capaz de atribuir um tom para a narrativa? (sarcasmo,

ironia ou até felicidade) e, como a máquina seria capaz de interpretar metáforas, por exemplo?

Assim como nos diz o linguista russo Roman Jakobson (1971, p.42):

“( . . . ) o significado de um signo linguístico não é mais que sua tradução por um outro signo que lhe pode ser substituído, especialmente um signo ‘no qual ele se ache desenvolvido de modo mais completo’, como insistentemente afirmou Peirce, o mais profundo investigador da essência dos signos. (...). Distinguimos três maneiras de interpretar um signo verbal: ele pode ser traduzido em outros signos da mesma língua, em outra língua, ou em outro sistema de símbolos não-verbais. ”

Podemos ver em nosso dia a dia o uso desse ainda falho aprendizado de máquina, retomando o conceito de que a T.A. não é, ainda, uma ciência exata e que o conjunto lexical “aprendido” por uma máquina pode ter vários significados, exemplificado na palma de nossas mãos quando usamos nossos celulares para enviar mensagens e estas são “corrigidas” (muitas vezes erroneamente) pelo auto corretor do próprio aplicativo que está sendo usado. O auto corretor “aprendeu” por amostragem e uso o conjunto lexical do usuário e irá “decidir” como usá-lo.

No caso da tradução de um idioma para outro, outros fatores além dos aspectos lexicais têm que ser levados em consideração como a sintaxe e a gramática. Para que faça sentido ao usuário da língua, o discurso tem que seguir uma sequência de regras para que a comunicação seja efetivamente feita. Porém, o não domínio dessas regras pode levar a máquina a não executar precisamente sua tarefa de tradução. (VIEIRA; LIMA, 2001)

Podemos ver este exemplo no excerto do livro da Professora Adelaide H. P. Silva (2019, p.108), onde temos:

“Considere como exemplo uma sentença do inglês: “*the teacher read a fairy tale to the students*”. Como traduzir essa sentença para o português? A primeira questão que se coloca é o gênero gramatical atribuído a “*the teacher*”. Como sabemos, podemos traduzir esse sintagma tanto como “o professor” quanto “a professora”. Como escolher? Para um humano, a resposta é simples e rápida: pelo contexto. Entretanto, para uma máquina, essa resposta não é suficiente. É preciso indicar qual é o contexto e como ele se relaciona ao excerto que queremos traduzir. É preciso, portanto, fazer a máquina “enxergar” as relações sintáticas envolvidas no processo e as relações lógico-formais que o embasam. ”

Retomando os conceitos de Aprendizado de Máquina e Tradução Automática, ferramentas foram criadas para tentar tornar o trabalho de traduzir mais eficaz, dinâmico e capaz de acompanhar a velocidade da comunicação contemporânea.

Surge então a ideia das *CAT Tools* (*Computer-aided Translation Tools*, ou Ferramentas de Tradução Auxiliadas por Computador) que têm sido estudadas e usadas desde os anos 1980, década frutífera para as tecnologias computacionais e a implementação global da internet e, conseqüentemente, a comunicação global e a necessidade de uma comunicação mais ágil.

As *CAT Tools* são programas de computadores, muitas vezes online e gratuitos, com a capacidade de armazenamento prévio de um glossário (preferencialmente de termos específicos para o projeto de tradução ou versão que será executado) inserido pelo usuário, uma *Translation Memory*, ou Memória de Tradução, doravante denominada de M.T., e a ajuda da Tradução Automática.

Este tripé de informações, organizadas dentro de uma *CAT Tool*, é capaz de tornar as traduções ou versões de documentos em um trabalho mais eficaz, mais uniforme e de mais fidedigno.

Existem *CAT Tools* que são gratuitamente disponibilizadas na internet e algumas que são pagas (licenças de uso e outros artifícios). Com o propósito ilustrativo, podemos citar duas ferramentas gratuitas e online que estão disponíveis na internet.

A primeira delas que podemos citar é o website *Smarcat* ([www.smartcat.com](http://www.smartcat.com)) que é uma ferramenta russa criada em 2016 e que conta com serviço para usuário freelancer e para trabalhos em equipes. Além de disponibilizar várias opções de Tradução de Máquina, ainda conta com um serviço de OCR (*Optical Character Recognition*, ou seja, uma ferramenta que reconhece caracteres dentro de uma imagem ou foto) possibilitando assim o trabalho com arquivos de imagem ou PDFs (*Portable Document Format*, ou, Documento em Formato Portátil), facilitando muito o trabalho do tradutor em um projeto.

Outra ferramenta italiana muito famosa criada em 2011 é o *Matecat* ([www.matecat.com](http://www.matecat.com)) que, além de trazer várias funcionalidades comuns às *CAT Tools*, ainda conta com um serviço muito prático chamado de *Aligner* (Alinhamento) que proporciona ao usuário-tradutor a experiência de usar uma tradução prévia feita dentro de

uma *CAT Tool* e transformá-la em uma Memória de Tradução, sendo assim inserida no programa para facilitar projetos de assuntos similares.

Não obstante, ainda se trata de uma ferramenta que, como dito ao longo deste estudo, não exclui o trabalho humano. Fatores como a historicidade, o contexto social, o tom, por exemplo, a ser usado, não podem ser ainda apresentados por uma máquina.

O glossário pode ser definido como o conjunto de termos selecionados pelo usuário que constará termos que, preferencialmente, terão a melhor tradução por definição encontradas em um dicionário. O tradutor deverá pesquisar e selecionar previamente, mediante o assunto abordado em dado projeto, e incluir a melhor (ou mais próxima) definição de cada termo específico àquele tema.

A M.T. (Memória de Tradução) é o armazenamento de palavras, trechos, frases ou segmentos, feitos anteriormente e armazenados nas *CAT Tools*. Essa memória, ou arquivamento de dados previamente instaurados, é acionada pelo usuário quando encontra similaridades linguísticas e, assim, serem utilizadas novamente. Se pensarmos em vários projetos de tradução que contém um tema comum, as M.T. poupam o trabalho do tradutor na re-pesquisa desses termos comuns.

Tanto na preparação e inserção de um glossário específico, como no uso das M.T.s, o campo da Linguística Computacional chamará esse arquivo lexical de Linguística de Corpus.

“ A Linguística de Corpus preocupa-se basicamente com o trabalho a partir de corpora eletrônicos que contenham amostras de linguagem natural. Essas amostras podem ser de diferentes fontes. Por isso, podemos encontrar os mais variados bancos de corpora eletrônicos: há corpora de linguagem falada, corpora de linguagem escrita literária, corpora com textos de jornal, corpora compostos exclusivamente por falas de crianças em estágio de desenvolvimento linguístico, etc. Os trabalhos envolvendo corpora linguísticos nem sempre têm como objetivo produzir algum software ou aplicativo. Normalmente, eles estão voltados para o estudo de determinados fenômenos linguísticos e sua ocorrência em grandes amostras de uma determinada língua (ou de uma variedade, dialeto ou modalidade dela) (OTHERO, 2006, p.342)

De posse desse *corpora linguístico* pré-acrescentado ao programa, resta a ativação de Tradução Automática. O mais famoso programa utilizado em Tradução Automática atualmente é a *Google Translate* (G.T.) onde uma tradução estatística acontece. A (s) palavra (s), termo(s) ou frase(s) são inseridas de um lado e a máquina escolhe a(s)

definição(ões) mais utilizados globalmente para gerar uma resposta satisfatória. “Obviamente, o sistema não chegará a oferecer 100% de precisão, porém, o produto final possibilita a compreensão do sentido do texto de origem. (AIKEN e BALAN, 2011 IN: NETO, 2013)

## CONCLUSÃO

Tendo em vista a trajetória aqui traçada da evolução da tradução e versão com o advento da evolução tecnológica e estando o tradutor munido desse conhecimento perante o letramento digital contemporâneo, o uso das *CAT Tools* só tem a agregar no trabalho dos tradutores em termos de produtividade e uniformidade de termos e corpora. Estando o tripé: Glossário, Memória de Tradução e Tradução Automática devidamente programados dentro da *CAT Tool* e, sendo seu acesso, online e gratuito, o trabalho do tradutor só tem a se tornar de mais qualidade e de mais fácil execução.

## REFERÊNCIAS

- ARAÚJO, K.; NOBRE, A. O uso inovador de objetos de aprendizagem sobre ferramentas informáticas de apoio à tradução escrita no âmbito do ensino e aprendizagem por via eletrônica para docentes em cursos superiores de línguas/tradução. 5º Seminário de Educação na Sociedade da Informação–Internet das Coisas, 2016.
- CERRI, R.; CARVALHO, ACPLF. Aprendizado de máquina: breve introdução e aplicações. *Cadernos de Ciência & Tecnologia*, v. 34, n. 3, p. 297-313, 2017.
- COSTA, G. C.; DE GÊNOVA DANIEL, F. Google Tradutor: análise de utilização e desempenho da ferramenta. *Tradterm*, v. 22, p. 327-361, 2013.
- ERWEN, Z.; WENMING, E. Application of Computer-Aided Translation Technology in Translation Teaching. *iJET – Volume 8, Issue 5, October 2013*.
- FIORIN, J. L. **Linguagem e interdisciplinaridade**. *Alea: estudos neolatinos*, v. 10, p. 29-53, 2008.
- JAKOBSON, R. **Aspectos linguísticos da tradução**. *Linguística e comunicação*, v. 15, p. 66-72, 1969.

MITCHELL, T.M. **Machine Learning** - McGraw-Hill Science/Engineering/Math; March 1, 1997.

NETO, F. F.S. Máquinas de tradução: ponto de vista histórico e prático. Artigo apresentado à Central de Cursos como requisito parcial para conclusão do curso de especialização Tradução Inglês, na Universidade Gama Filho – UGF.Brasília,2013.

OTHERO, G.A. Linguística Computacional: uma breve introdução. Letras de Hoje. Porto Alegre. v. 41, nº 2, p. 341-351, junho, 2006.

REIS, D.; MILTON, J. **História da tradução no Brasil**: percursos seculares. TRANSLATIO Porto Alegre, n. 12, dezembro de 2016.

RODRIGUES, R. S. V. Saussure e a definição da língua como objeto de estudos. ReVEL. Edição especial n. 2, 2008. ISSN 1678-8931 [www.revel.inf.br].

RUSSEL, S.J. **Inteligência Artificial**. Elsevier. Rio de Janeiro, 2013.

SILVA, A. H. P. **Introdução à Linguística**. - 1. ed. - Curitiba [PR] : IESDE, 2019.

VIEIRA, R.; LIMA, V. L. S. ERI/Linguística computacional: princípios e aplicações. In: Luciana Porcher Nedel. (Org.). IX Escola Regional de Informática - 2001 (ERI2001). 1 ed. Porto Alegre: SBC- Regional Sul, 2001, v. 1, p. 27-58.

*Recebido em: 03/02/2022*

*Aprovado em: 05/03/2022*

*Publicado em: 10/03/2022*